

学校编码: 10384

分类号_____密级 _____

学 号: 200231048

UDC_____

厦 门 大 学
硕 士 学 位 论 文

基于商品分类的电子商务推荐系统的
研究与设计

The Research and Design of E-commerce
Recommendation System Based on Product Taxonomy

曾 潮 缤

指导教师姓名: 罗 键 教授

专 业 名 称: 系 统 工 程

论文提交日期: 2005 年 5 月

论文答辩日期: 2005 年 6 月

学位授予日期: 2005 年 月

答辩委员会主席: _____

评 阅 人: _____

2005 年 6 月

厦门大学学位论文原创性声明

兹提交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文而产生的权利和责任。

声明人（签名）：

年 月 日

摘 要

随着 Internet 迅速发展，电子商务网站越来越受到重视。作为企业对外经营的场所，电子商务系统在为用户提供越来越多的选择的同时，结构也变得越来越复杂。用户在面对大量的商品空间时，很难快速有效的找到自己喜爱的商品。电子商务推荐系统直接与用户进行交互，模拟商店销售人员向用户提供商品推荐，从而帮助用户完成商品购买。电子商务推荐系统在有效提高网站产品的吸引力、转化网站的浏览者为购买者、建立和增强顾客的忠诚度、增加网站的交叉销售量方面起着重要的作用。协作过滤技术是电子商务推荐系统中应用最为成功的个性化推荐技术之一，但是，当系统中的用户数目和商品的数目达到很大规模时，协作过滤技术就将面临很大的挑战。为了解决这种问题提高推荐系统的推荐质量和实时性，国内外的研究人员进行了大量的研究，并且已经取得了一定的成果。

论文全面介绍电子商务推荐系统的概念、分类及使用的技术。在此基础上详细阐述了协作过滤技术及已有的两个协作过滤算法。文中着重分析了协作过滤算法在推荐系统中面临的主要问题，提出了根据网站的商品分类建立商品分类集合，并且根据这种分类集合对高度稀疏的用户兴趣度矩阵进行划分，从而提高协作过滤算法的效率和推荐质量。然后给出了算法在标准数据集上的实验结果，并对改进的算法和原算法的相关性能的比较分析，证明了算法的有效性和可行性。文中还结合改进的算法和现有的技术设计和实现了一个原型系统。最后，对本文的研究进行了全面的总结，指出了进一步的研究方向。

关键词：电子商务 推荐系统 协作过滤 算法 商品分类

ABSTRACT

With the rapid development of Internet, e-commerce website becomes more and more popular. The structure of the e-commerce website is getting complex than ever if aiming to provide growing various products to customers. This situation makes it hard for customers to find the products they really prefer. Recommendation systems interact directly with customers, and recommend products to customers like a salesman. Recommendation systems can enhance e-commerce sales by making the website products much more attractive, converting browsers into buyers, increasing cross-sell and building loyalty to prevent user losing. Collaborative filtering is kind of successful technology that is implemented in recommendation systems nowadays. But with the on-going increasing of users or products in the systems, collaborative filtering systems are facing great challenge. In order to resolve the problem and improve the quality and efficiency of recommendation systems, many researches have been carried out all over the world and achieved a lot.

The paper introduces the concept, classification and use of e-commerce recommendation systems. Based on these, the paper elaborates collaborative filtering technology and its two directions of algorithms in detail. The paper takes great effort to analyze the problems in current collaborative filtering technology; we propose a new collaborative filtering algorithm based on product taxonomy. With the algorithm, we can improve the quality and efficiency of recommendation system. Then the experimental results deprived from the benchmark experimental dataset prove the validity and feasibility of the new algorithm. Finally, it summarizes the weak points and researching directions for the future.

Key Words: E-commerce; Recommendation System; Collaborative Filtering; Algorithm; Product Taxonomy

目 录

第一章 绪 论	1
§1.1 论文背景及其研究内容	1
§1.2 论文的组织结构	2
第二章 电子商务推荐系统及其关键技术	4
§2.1 电子商务推荐系统	4
§2.1.1 电子商务推荐系统的作用	5
§2.1.2 电子商务推荐系统的构成	6
§2.1.3 电子商务推荐系统分类	8
§2.2 电子商务推荐系统关键技术	11
§2.2.1 数据挖掘	11
§2.2.2 Web 使用挖掘	14
§2.2.3 协作过滤技术	19
§2.3 小结	20
第三章 协作过滤技术综述	21
§3.1 协作过滤技术简介	21
§3.2 协作过滤的实现	23
§3.3 协作过滤算法	28
§3.3.1 全局数值算法	28
§3.3.2 基于模型的算法	29
§3.3.3 协作过滤算法的局限性	30
§3.4 小结	31
第四章 基于商品分类的协作过滤算法	32
§4.1 基于商品分类的协作过滤算法的实现过程	32
§4.1.1 商品分类集合的建立	33

CONTENTS

Chapter I Introduction	1
§ 1.1 The Background and Research Context of The Paper	1
§ 1.2 The Structure of The Paper.....	2
Chapter II E-commerce Recommendation System & Its Key Technologies.....	4
§ 2.1 E-commerce Recommendation System.....	4
§ 2.1.1 The Function of E-commerce Recommendation System.....	5
§ 2.1.2 The Composition of E-commerce Recommendation System ..	6
§ 2.1.3 The Classification of E-commerce Recommendation System.	8
§ 2.2 The Key Technologies of E-commerce Recommendation System.	11
§ 2.2.1 Data Mining	11
§ 2.2.2 Web Usage Mining.....	14
§ 2.2.3 Collaborative Filtering.....	19
§ 2.3 Summary.....	20
Chapter III A Summary Introduction for Collaborative Filtering Technology	21
§ 3.1 A Brief Introduction for Collaborative Filtering Technology	21
§ 3.2 The Process of Collaborative Filtering Technology	23
§ 3.3 Collaborative Filtering Algorithm.....	28
§ 3.3.1 Memory-Based Algorithm	28
§ 3.3.2 Model-Based Algorithm	29
§ 3.3.3 The limitation of Collaborative Filtering Algorithm.....	30
§ 3.4 Summary.....	31

Chapter IV A Collaborative Filtering Algorithm Based on Product Taxonomy	32
§ 4.1 The Process of Collaborative Filtering Algorithm Based on Product Taxonomy	32
§ 4.1.1 The Definition of Classification Set	33
§ 4.1.2 Customer Profile Creation	38
§ 4.1.3 Neighborhood Formation	40
§ 4.1.4 Recommendation Generation	42
§ 4.2 Experimental Result & Analysis	43
§ 4.2.1 Experimental Dataset	43
§ 4.2.2 The Measurement of Experiment	44
§ 4.2.3 Experiment Results	46
§ 4.3 Summary	46
Chapter V System Design & Implementation	48
§ 5.1 Technology Architecture	48
§ 5.2 System Design	50
§ 5.3 System Running	52
§ 5.4 Summary	54
Chapter VI Conclusion & Perspective	55
§ 6.1 Conclusion	55
§ 6.2 The Directions for Further Research	55
References	57
Thanks	61

第一章 绪 论

§ 1.1 论文背景及其研究内容

Internet 的迅速发展,将人类带入信息时代和网络社会,对我们的生活产生了深远的影响。Internet 的发展正在极大的改变我们的生活,网上购物使得我们不出家门就可以随心所欲的选择自己喜欢的商品。电子商务正以令人难以置信的速度蓬勃发展,新的商业环境为企业提供商机的同时,也对企业提出了挑战。由于电子商务的虚拟环境下,商家在网上提供的商品的种类和数目繁多,用户很难快速有效的查找到自己喜欢的商品。在这种背景下,用户就急需电子商务系统能够提供一种类似购物助手的功能来帮助选择商品,它根据用户的兴趣和爱好自动推荐给每个用户其可能感兴趣和满意的商品。这样做的根本目的是使得用户在虚拟环境下的购物更加方便、更加容易。近年来兴起的推荐系统成为解决这些问题的重要途径之一。

推荐系统(Recommendation System)就是根据用户的个人的喜欢、习惯来向其推荐商品和信息的程序。早期的推荐系统提供的推荐信息是对全体用户提供相同的信息,如商品销售排行版等,并不是对个体用户进行推荐,这类推荐系统在个性化方面存在不足。随着电子商务的发展,为适应实际商务中客户为得到一对一服务的需要,应用于电子商务网站的推荐系统逐渐向能实时地为客户提供符合其个人需要的推荐服务的方向发展,从而提高客户对自身网站的信任度,为企业的数字化发展和企业经济效益的增长提供更广泛的发展前景。因此加强对电子商务推荐系统的研究具有非常重要的现实意义。

协作过滤推荐是至今最成功的个性化推荐技术,被广泛的应用到多个领域中,它分析“用户—用户”关系,根据用户兴趣相似性来推荐资源,把和用户相似的其他用户的意见推荐给特定的用户。无需考虑资源的表示

形式，把用户对资源的评分向量作为用户兴趣的表示形式。它能够过滤难以机器内容自动分析的信息，象艺术品、音乐、电影等；能够基于一些复杂的，难以表达的概念(如质量、品味)进行过滤；具有推荐新信息的能力。但是也面临两方面的挑战：(1)数据的高维稀疏性。(2)算法的可扩展性不好。

针对上述提到的推荐系统重要性及协作过滤算法存在的问题，本文参考了国内外研究人员的最新的研究成果，结合自己的理论和实践的基础，在以下几个方面做了一些研究：

1) 总结电子商务推荐系统的概念、作用和分类等，介绍了数据挖掘和Web使用挖掘技术及协作过滤技术。详细介绍了协作过滤技术及其在电子商务推荐系统中的应用和所面临的问题。

2) 根据电子商务的特点给出了根据网站的商品分类建立商品分类集合的方法及分类集合调整的算法，并且提出了根据建立的分类集合对协作过滤算法中的用户兴趣度矩阵进行划分，以解决数据高维稀疏性和算法的扩展性问题，同时可以改善推荐系统推荐的准确率和覆盖率。

3) 利用一个被广泛使用和认可的实验数据集对改进算法进行实验验证，并对改进算法和原算法进行性能对比与评价

4) 在改进算法的基础上，给出推荐系统的设计思想及原型系统的实现。

§ 1.2 论文的组织结构

本论文的组织结构安排如下：

第二章对电子商务推荐系统的基本概念、获取信息的方式、推荐系统的分类进行了概述，在此基础上介绍了实现电子商务推荐系统的关键技术，主要介绍了数据挖掘、Web 使用挖掘技术。

第三章对协作过滤技术进行了全面的介绍，重点介绍了协作过滤技术的实现步骤、算法及其优缺点。

第四章针对协作过滤技术存在的问题，根据商品分类的模型，给出了商

品分类集合的划分和调整方法，并且在这个方法的基础上对协作过滤算法进行改进，并用实验的方法验证了该方法的有效性。

第五章是对基于商品分类协作过滤算法的实现，给出了系统的设计思想、总体结构，及一个原型系统的实现。

第六章对论文进行总结，给出了进一步研究的方向。

第二章 电子商务推荐系统及其关键技术

电子商务是组织或个人用户在以 Internet 为基础的计算机系统支持下所进行的商务活动。Internet 的迅速发展带动了电子商务在全球的发展,然而随着网上信息和用户的迅速增长,于是带来了网站和用户两个方面的问题:一方面,大量网站的不断涌现,使得各网站对网上业务的争夺变的异常的激烈,因而网站业务成败的关键之一在于网站所提供的服务的质量,他们要考虑如何为合法的用户,在正确的时间里方便的提供正确的信息,以吸引更多的用户;而另一方面,用户在应用网络时面对铺天盖地的信息,却常常因找不到自己所需要的内容变的无所适从。70%的 Web 用户认为现有的网站难以提供有效的帮助^[11]。

传统的电子商务网站采用的是“一对多”的批发方式,即对所有的用户提供统一的界面、同样的内容,而新的竞争要求采用“一对一”的零售方式,及针对不同的顾客或顾客群提供不同的服务形式和内容。例如,一个网上书店的顾客,由于职业不同、年龄不同、喜好不同等,他们所关心的内容可能是完全不同。如同一个老练推销员会对不同的顾客采用不同的推销策略来推销产品一样,一个成功的网站也应该具备这种针对性的智能,即改变“一对多”的批发模式为“一对一”的零售模式。电子商务个性化推荐系统(Recommendation System)正是针对这一问题和要求应运而生^[12]。

§ 2.1 电子商务推荐系统

电子商务推荐系统就是利用统计和知识发现技术在电子商务系统中向顾客提供商品信息和建议,帮助顾客决定购买何种商品,模拟销售人员向客户推荐商品完成商品的购买过程。推荐系统推荐何种商品是在电子商务网站整体商品的购买情况、客户的人数统计或者对客户购买的历史记录上

进行分析产生的。从广义上讲这些技术带有个性化的色彩,因为对于不同的客户系统表现出一定的自适应性。推荐系统自动产生个性化,为每个客户提供个性化的服务,个性化技术是实现推荐系统的关键。Amazon的CEO感慨的谈到“如果我们有两百万个用户在网上,我们就应该建立两百万个商店”。^[13]目前,推荐系统已广泛运用到各行业中,推荐对象包括书籍、音像、网页、文章和新闻等(如图2-1)。

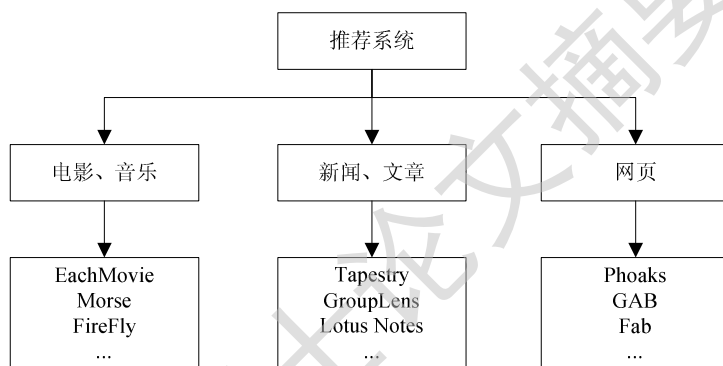


图2-1 主要的推荐系统

§ 2.1.1 电子商务推荐系统的作用

电子商务推荐系统最大的优点在于它能收集用户感兴趣的资料,并根据用户兴趣偏好主动为用户做出个性化推荐。这样,当用户每次输入用户名和密码登录电子商务网站后,推荐系统就会自动按照用户偏好程度的高低推荐给用户最喜爱的N个产品,而且给出的推荐是实时更新的。总的来说,推荐系统在电子商务网站的主要作用表现为以下三个方面:

1) 大大增加网站的客户,转化网站的浏览者为购买者。访问站点的经常是只看不买,出现这种情况很多时候是因为站点的设计使他们不能便捷地找到他们感兴趣或想购买地物品。而推荐系统就能帮助他们便捷地找到想购买地商品,从访问者变成购买者。

2) 增加交叉销售量。推荐系统通过向客户推荐额外的商品来提高站点企业地交叉销售量。例如站点可以根据客户当前购物车中的物品向他们推荐一些和这些已选购的物品相关的物品。如果有一个比较好的推荐系统,

则企业的平均定购量就可能增加。

3) 建立和加强顾客忠诚度。在电子商务环境下, 站点客户要到竞争者那里只需要简单的点击操作, 相对传统商务形式来说, 建立和增加客户的忠诚度显得更为重要。站点对客户和一般访问者不断得了解和学习, 利用推荐系统提供符合客户所需得Web界面, 客户则通过不断的访问这个最符合他们各自所需的站点来回报站点企业。越多的用户得到站点推荐系统给他们提供的个性化服务, 则他们会对站点赋予更多的忠诚。

研究表明, 在基于电子商务的销售行业使用个性化推荐系统后, 能提高销售额2%-8%^[3], 尤其在书籍、电影、CD音像、日用百货等产品相对较为低廉且商品种类繁多的行业, 推荐系统能大大提高企业的销售额。

§ 2.1.2 电子商务推荐系统的构成

整个电子商务推荐系统的组成主要可以分为三个模块: 输入功能(Input functional) 模块、输出功能 (Output functional) 模块、推荐方法 (Recommendation method)模块。

1. 电子商务推荐系统的输入

不同类型的电子商务推荐系统, 其输入信息也不相同。不同电子商务推荐系统根据不同的输入信息产生不同类型的推荐, 电子商务推荐系统的输入包括多种形式, 主要包括^[14]:

1) 隐式输入: 将用户访问电子商务Web站点的浏览行为作为推荐系统的输入, 用户的浏览行为与访问一般的Web站点没有区别。

2) 显式输入: 也是将用户的浏览行为作为电子商务推荐系统的输入, 但与隐式输入不同, 用户的显式浏览输入是有目的的向电子商务推荐系统提供自己的兴趣爱好。

3) 关键词/商品属性输入: 用户在搜索引擎中输入关键字作为推荐系统的输入, 或者将用户当前正在浏览的商品类别作为推荐系统的输入。

4) 用户评分输入：将用户对商品的数值评分数据作为推荐系统的输入。电子商务推荐系统列出一系列商品让用户评分，用户的评分可以是一个数值，数值大小表示用户对商品的喜好程度。

5) 用户文本评价输入：用户对已经购买的商品或自己熟悉的商品以文本的形式进行个人评价，推荐系统本身并不能判断这些评价的好坏。其他用户浏览该商品时，可以看到用户对商品的文本评价信息。

6) 编辑推荐输入：将领域专家对特定商品的评价作为推荐系统的输入，领域专家对商品的性能特点进行全面详细的介绍，用户通过专家的专业介绍，可以对自己并不熟悉的商品加深认识，从而决定是否购买该商品。

7) 用户购买历史输入：推荐系统将用户的购买历史作为隐式评分数据。一旦用户购买了特定商品，则认为用户喜欢该商品。推荐系统根据用户的购买历史产生相应的推荐。但是用户购买了某件商品并不代表用户喜欢该商品，所以在精确的推荐系统中，用户可以对购买的商品进行重新评分，从而使推荐系统产生更精确的推荐。

2. 电子商务推荐系统的输出

不同类型的电子商务推荐系统，其输出也各不相同。大型电子商务系统可以同时向用户产生多种不同形式的输出。电子商务推荐系统的输出形式主要包括^[14]：

1) 相关商品输出：推荐系统根据用户表现出来的行为特征或电子商务系统的销售情况向用户产生商品推荐，这种方式是电子商务推荐系统中最普遍的一种输出。相关商品输出可以基于简单的销售排行向用户推荐热门商品；也可以基于对用户的行为特征进行深入分析，发现用户的购买行为模式，从而产生个性化的推荐。

2) 个体文本评价输出：电子商务推荐系统向目标用户提供其他用户对商品的文本评价信息个体文本评价一般是非个性化的，对每个项而言，所有用户得到的个体文本评价均相同。

3) 个体评分输出：向目标用户提供其他用户对商品的数值评分信息。个体评分输出没有大量的文本描述信息，因此更加简洁明了。个体评分输出比较适合于个体数值评分数据比较少的场合。

4) 平均数值评分输出：电子商务推荐系统向用户提供其他用户对商品数值评分信息的平均值。这种输出形式具有简洁明了的优点，用户可以立即获得对该商品的总体评价。

5) 电子邮件输出：电子商务推荐系统通过电子邮件的形式向用户提供商品的最新信息。这种输出形式可以吸引用户再次访问电子商务系统，从而达到保留用户，防止用户流失的目的。

6) 编辑推荐输出：向用户提供领域专家对商品的专业介绍，用户通过专家的专业介绍可以对自己并不熟悉的商品加深认识，了解商品的性能特点，从而决定是否购买该商品。

3. 推荐方法模块

推荐方法模块是推荐系统的核心部分，负责由输入如何得到输出，决定着推荐系统的性能优劣。推荐方法模块以推荐技术和算法为技术支撑，具体的推荐技术将在本章后面部分详细介绍。

§ 2.1.3 电子商务推荐系统分类

电子商务推荐系统向客户进行推荐的方式有许多种，它可以是建议或者通过计算得到的预测，也可以是其他客户对产品的个人评价、评论等，而选择哪一种方式主要取决于该电子商务网站希望客户如何使用推荐。

1. 推荐系统的界面表现形式分类^[37]

按照推荐系统的界面表现形式主要分为以下几种：

1) 浏览(Browsing)：客户提出对特定商品的查询要求，推荐系统根据查询要求返回高质量的推荐。

2) 相似项(Similar Item)：推荐系统根据客户购物篮中的商品或客户感

兴趣的商品推荐类似的商品，为客户提供个性化的推荐。

3) 电子邮件(Email): 推荐系统通过电子邮件的方式通知客户可能感兴趣的商品信息, 使网站与客户保持联系, 提高客户对网站的信任度, 从而使增加对该网站的访问量。

4) 评论信息(Text Comments): 推荐系统向客户提供其他客户对相应产品的评论信息客户根据他人对产品好坏的评价, 来做出自己的判断。

5) 等级评价(Average Rating): 推荐系统向客户提供其他客户对相应产品的等级评价, 而不是产品的评论信息, 通过对等级评价的相应统计和分析, 较直观地表示出其他客户对产品的观点或看法, 使客户易于接受该种推荐。

6) Top-N: 推荐系统根据客户的喜好向客户推荐最可能吸引他的 N 件产品, 一方面可以把网站的浏览者转变为客户, 另一方面帮助客户决定是否购买自己最初感到犹豫不决的产品。

7) Ordered Search Results: 推荐系统列出所有的搜索结果, 并将搜索结果按照客户的兴趣度降序排列。

2. 推荐系统使用技术分类

按推荐系统使用技术分类, 即是从客户的角度出发, 按照技术的自动化程度(Degree of Automation)和持久性程度(Degree of Persistence)两个标准进行的分类。

1) 自动化程度是指客户为了得到推荐系统的推荐是否需要显式地通过手工或自动地输入信息的程度。自动方式意味着推荐的产生不需要客户明显的任何操作, 例如当客户与网站进行交互时, 恰好出现客户感兴趣的推荐。手工方式意味着客户得到符合自己兴趣的推荐前需要进行一些显式操作如对产品进行评价等。

2) 持久性程度指推荐系统产生的推荐是基于客户当前的单个会话(Session)还是基于客户的多个会话, 是瞬时还是持久。瞬时推荐只考虑当前

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库